

Implementation of RSS Feed Reader

Phyu Hnin Thwe, Dr. Khin Nwe Ni Tun
phyuhninthwe88@gmail.com
University of Computer Studies, Yangon

Abstract

RSS (Really Simple Syndication) is a technology that is being used by millions of web users around the world to keep track of their favourite websites. By using RSS, we can stay on top of the news and information without repeatedly checking multiple web sites to see if they have been updated. RSS is a time-saving way to receive and information updates (often called “RSS feeds”, “news feeds” or simply “feeds” from favourite Web sites and blogs. To read RSS feeds, an RSS reader (also called a “news aggregator” or a “feed reader” or “RSS aggregator”) is needed. RSS reader is a piece of software that checks RSS feeds and read any new articles.

In this paper, design and implementation of the RSS aggregator is discussed. Testing results with C# implementation of windows based program of RSS reader is described. Input data of this system can be used from RSS feed from online pages or RSS feed files for test cases.

Keywords: *RSS, Really Simple Syndication, news feeds, blogs, feed reader, RSS aggregator, C#, RSS feed files.*

1. Introduction

With the rise of always-on Internet connections in homes and businesses, and the continued explosive growth of the World Wide Web and Internet-accessible applications, it is becoming more and more important for applications to be able to share data with each other. Sharing data among disparate platforms requires a platform-neutral data format that can be easily transmitted via standard Internet protocols—this is where XML fits in. Since XML files are essentially, simple text files with well-known encodings and since there exist XML parsers for all commonly used programming languages, XML data can be easily consumed by any platform.

A good example of data-sharing using XML is Web site syndication, commonly found in news sites and Web logs. With Web site syndication; RSS, a Web site publishes its latest content in an XML-formatted, Web-accessible syndication file.

The basic idea of restructuring information about web sites goes back to as early as 1995, when Ramanathan V. Guha and others in Apple Computer's Advanced Technology Group developed the Meta Content Framework (MCF). Today many different types of content are syndicated on the internet. Millions of online publishers, including newspapers, commercial websites and blogs, now publish their latest news headlines, product offers or blog postings in standard format news feed [8]. Web syndication cut down many tasks such as: Convenience and Simplicity (difference between text vs. graphical browsers), Connected in real time and anytime and Personal Learning Environment [9].

RSS stands for Really Simple Syndication or *Rich Site Summary*. It is used for sharing web contents in the form of xml dynamically. RSS is read by various free software's called as “RSS Reader”. If dynamic web contents or the contents which are regularly changes in website, it is good to use RSS functionality. Some of the main advantage of using RSS is: (1) It reduces traffic in website. (2) It saves time of the user because now the user needs not to visit your website each time. (3) It automatically updates the user about website updated contents.

There are a number of syndication formats in use, one of the more popular ones being RSS 2.0. (RSS 2.0 Specification is published online at the Technology at Harvard Law site.) Once a Web site has publicly published a syndication file, various clients may decide to consume it. There are a number of ways to consume a syndication file. Syndication files are also commonly consumed by news aggregator applications, which are applications designed specifically to retrieve and display syndication files from a variety of sources. This paper discusses the structure of RSS 2.0 feed format, extracts updated feeds according to the user choices.

2. RSS Feed Reader

The Internet is full of dynamic and engaging content. The problem of how to reliably make content available for easy and timely access has been addressed in the form of web feeds. Web feeds allow content creators to put out new content quickly and easily. Web feeds are accessed via applications or services called aggregators. An aggregator finds web feeds and interprets them for the user in a way that

allows the user to access the content described by the web feed.

RSS is created using XML or eXtensible Markup Language, which is a markup language similar to HTML. All fields are defined. Tags are used to denote the field's classification. RSS is a family of XML file formats for Web Syndication used by (amongst other things) news websites and weblogs. The abbreviation is used to refer to the following standards:

- Rich Site Summary (RSS 0.91)
- RDF Site Summary (RSS 0.9 and 1.0)
- Really Simple Syndication (RSS 2.0)

The RSS formats provide web content or summaries of web content together with links to the full versions of the content, and other meta-data. This information is delivered as an XML file called RSS feed, webfeed, RSS stream, or RSS channel. RSS makes it possible for people to keep up with their favorite web sites in an automated manner that's easier than checking them manually.

Like HTML, proper construction requires that tags are both opened and closed. Example: `<title> Title of Item in Feed </title>` RSS has been around for more than a decade, but only recently the standard has been embraced by bloggers, webmasters and large news portals as a means of distributing information, in a standardized format. The first line in XML declaration defines the XML version and the character encoding used in the document. In this case the document conforms to the 1.0 specification of XML and uses the ISO-8859-1 (Latin-1/West European) character set. The next line is the RSS declaration which identifies that this is an RSS document (in this case, RSS version 2.0).

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<rss version="2.0">
<channel>
<title>W3Schools Home Page</title>
<link>http://www.w3schools.com</link>
<description>Free web building
tutorials</description>
<item>
<title>RSS Tutorial</title>
<link>http://www.w3schools.com/rss</link>
<description>New RSS tutorial on
W3Schools</description>
</item>
<item>
<title>XML Tutorial</title>
<link>http://www.w3schools.com/xml</link>
<description>New XML tutorial on
W3Schools</description>
</item>
</channel>
</rss>

```

Figure 2.1 Simple RSS Document

The next line contains the `<channel>` element. This element is used to describe the RSS feed [7]. RSS documents use a self-describing and simple syntax. Figure 2.1 shows a simple RSS document. Figure 2.2 depicts a configuration of RSS Feed Aggregator or Reader, it shows how the websites, the RSS feed XML files, and personal computer are connected. In this figure, a web browser being used to read first Web Site 1 over the Internet and then Web Site 2. It also shows the RSS feed XML files for both websites being monitored simultaneously by an RSS Feed Aggregator.

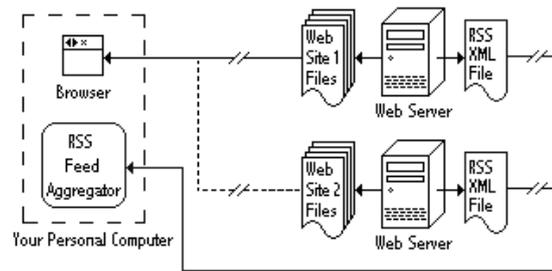


Figure 2.2 RSS Feed Aggregator

An RSS feed aggregator is a service that monitors large numbers of feeds. It allows users to subscribe to the content that they are interested in without explicitly specifying which RSS feeds the content is coming from. This is particularly convenient for the user, since the number of RSS feeds that can carry information of interest to the user can be very large. RSS feed aggregators use pull-based architectures, where the aggregator pulls RSS feeds from a web site that hosts the feed as shown in Figure 2.3.

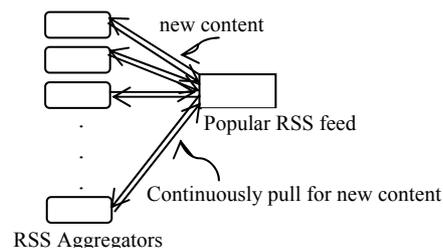


Figure 2.3 Pull-based method of RSS Aggregator

3. Related Work

With this growing emphasis on XML data, being able to work with XML data in an ASP.NET Web page is more pertinent now than ever before. Since Web site syndication is becoming all the rage, in [2] online news aggregator application is discussed. In [5], authors described and depicted a web browser being used to read first Web Site 1 over the Internet and then Web Site 2. It also shows the RSS feed

XML files for both websites being monitored simultaneously by an RSS Feed Aggregator. In [1], author described Using RSS and Weblogs for E-Learning which takes facilitating communication across project teams during design and development or between instructors and learners when they are separated by time and space, or between learners while building a community. [5] discussed an overview about FeedReader product portfolio and describe functions and architecture of Newsbrain engine. Newsbrain engine can be used in applications with user interface or in system services running without user interface. In [3], authors described CMS-ToPSS, a novel extension to content management systems, for scalable dissemination of RSS documents, based on the publish/subscribe model.

4. Design and Implementation

This session discuss the design and implementation of the system. In this system, once subscribed to a feed, an aggregator is able to check for new content at user-determined intervals and retrieve the update. By adding a feed reader, a user becomes subscribed to the feed. The basic way to subscribe is by simply clicking on the RSS icon and /or text link and then copy link location and pasting the resulting link directly into a feed reader. Figure 4.1 shows the system overview of the system.

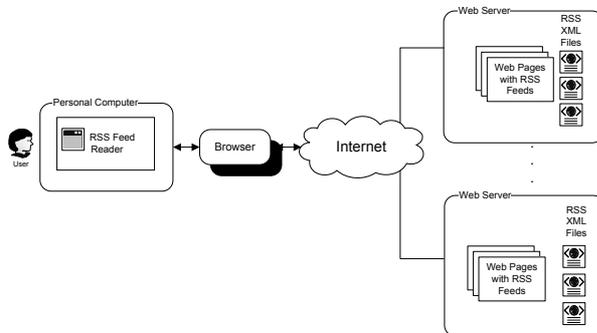
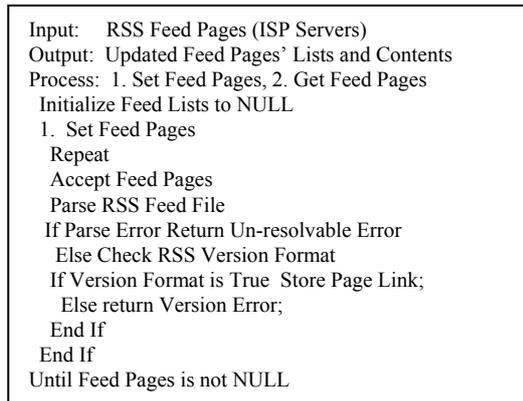


Figure 4.1 RSS Feed Reader System Overview

As this system is desktop type feed reader, users can subscribe to a feed by enter the feed's link into the reader or by clicking an RSS icon in a browser that initiates the subscribing process. The reader checks the user's subscribed feeds regularly for new content, downloading any updates that it finds. The technology of RSS allows internet users to subscribe to websites that have provided RSS feeds; these are typically sites that change or add content regularly.

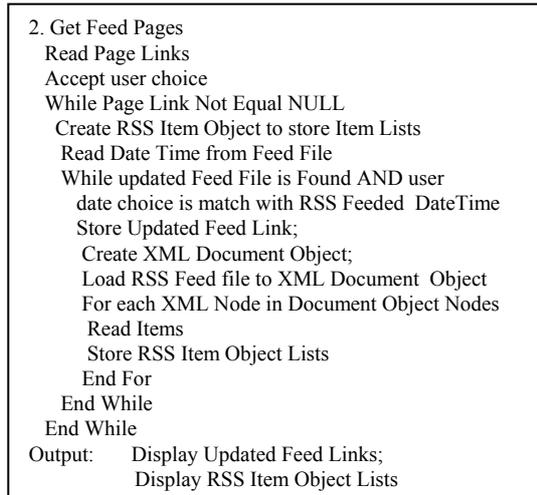
4.1 Algorithm of the Proposed System

By using RSS reader application, hundreds of web sites can be easily tracked every day because of updated information on many web sites can be browsed at a glance. This system analyzes the structure of RSS 2.0 feed format, extracts updated feeds according to the user choices. The basic algorithm for the RSS Feed Reader can be seen as shown in Figure 4.2.



(a)

In Figure 4.2 (a) the algorithm for setting feeds from web page's link are shown. Firstly, users has to input the feed link to the system and then the system saves the feed link's url and title for displaying the updated feed contents. In Figure 4.2 (b) the algorithm for getting feeds are shown. The input feed links and titles are get from the storage file then they have to be read by the reading processes with respect to the user options based on date filtering or browsing normally.



(b)

Figure 4.2 Feed Reader Algorithm (a) Setting Feeds (b) Getting Feeds

4.2 Sequence Diagram of the Proposed System

The proposed system is composed of four classes. They are RSS, RSSManager, ImportFeedForm (for capturing the feed links and titles) and MyRSSReader. (for displaying feed contents) RSS and RSSManger classes that encapsulate various aspects of the RSS feed aggregations and enable to associate with user interface classes. Figure 4.3 shows the sequence diagram of the system. The user browses the web sites and saves RSS feeded URL. After getting it, the system saves feed titles and urls. If the users want to view the feed titles, this system shows the titles from the saved file at any time. According to the user choices, the RSS feeded XML file is loaded to the DOM and the contents of feeds are parsed by it. If the RSS version is 2.0 this system can perform the operation steps, otherwise the system will give the message that cannot resolve the feed.

Once it parsed the operation steps, the contents of feed version, title, link, description and date are stored in the collection objects of this system. The updated feed contents are displayed according to the feed titles by user choices. Extracting of dated feeds are available in this system. This system can also show the related post by clicking the related urls of the feed contents until the user choose the next posts.

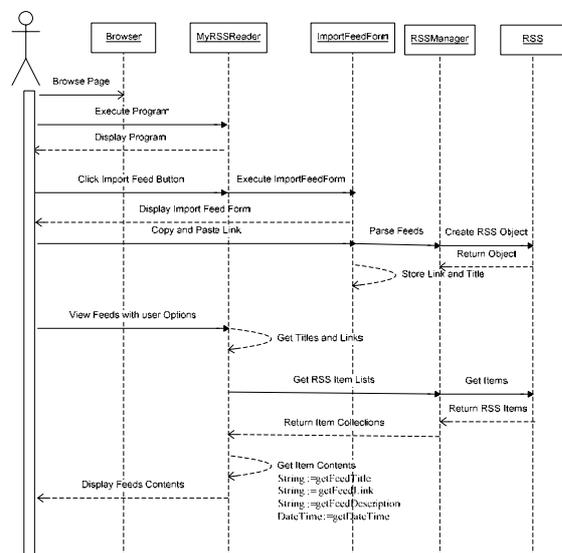


Figure 4.3 Sequence Diagram of the System

4.3 Testing Results

This system can work by scanning the worldwide web with latest postings based on the RSS code (containing the website's URL) provided or added by the user. When it finds a new posting, news, or update, it will publish the RSS feed on your home page containing the title of the posting, which also serves as a clickable link to the website source. This RSS feed may or may not contain the whole article, a summary, and photos, depending on what RSS aggregator you are using. This system is desktop-type reader in order to use free ware version. This system has also benefits of

- user friendly
- easy addition of RSS feeds
- can read updated feeds
- allows filtering of RSS feeds by date
- can extend the feed format for future use without special configuration
- easy navigation with all major browsers supported (e.g. Netscape Navigator, Internet Explorer, Firefox)

5. Conclusion

RSS Reader saves times, and it is proactive as it alerts the user when new information is available. The key benefit for the business is that they only have to produce and publish your information one time. It is extremely powerful to then have this information syndicated. This allows the business to save time and money. Distributing news, newsletter and announcements via RSS is much more cost effective than email. As RSS is becoming more widely used, RSS feeds are quickly making their way into every aspect of the Internet. RSS will continue to play a central role in determining how that information is syndicated and disseminated in powerful and useful ways. This system can transfer information from one website to another without any human intervention. This can also reduce the needed time to visit many individual web sites. Subscribing to a site's RSS feed and generated feeds can be read by this system.

References

- [1] B. Brandon, "Using RSS and Weblogs for E-Learning: An Overview", May 19, 2003.
- [2] C. De Brún, "RSS Workshop", Power Point Presentation.

- [3] M. Petrovic, Haifeng Liu, Hans-Arno Jacobsen, "CMS-ToPSS: Efficient Dissemination of RSS Documents", University of Toronto, 2005.
- [4] M. Scott, "Creating an Online RSS News Aggregator with ASP.NET", White Paper, Revised August 2003.
- [5] White Paper, "FeedReader RSS Platform".
- [6] "What is RSS?" A tutorial introduction to feeds and aggregators, Copyright (c) 2004 Software Garden, Inc.
<http://rss.softwaregarden.com/aboutrss.html>
- [7] "RSS Syntax",
http://www.w3schools.com/rss/rss_syntax.asp
- [8] http://en.wikipedia.org/wiki/Web_syndication
- [9] <http://www.w3.org/people/raggett/tidy/>